10/580032
IAP12 Rec'd PCT/PTO 19 MAY 2006

SPECIFICATION

METHOD OF ANALYZING A PROTEIN

TECHNICAL FIELD

The present invention relates to a method of analyzing a protein.

BACKGROUND ART

Every cell in eukaryote has essentially the same genetic information, and the genome information concerning human as well as many organisms are now being elucidated. Recently, it has also been found that a core structure called chromatin composed of a DNA and proteins is important for regulation of expression of the genome information. Proteins that bind to the DNA and form the core structure include histone H4, histone H2A, and histone H2B; and in addition, histone H1 is used for connecting chromatins.

In particular, the N-terminal side of each histone in chromatin is called "tail" because it protrudes out of the chromatin core structure,; and the site of modification, such as methylation, acetylation, and ubiquitylation or phosphorylation, on arginine, lysine, serine, or threonine in the region is responsible for regulation of the expression of particular genes and determines the epigenetic control mechanism during development and differentiation. Additionally, the regulation of gene expression, for

example, is related to various diseases including carcinogenesis, leukemia, aging, and others.

Accordingly, if it is possible to determine presence or absence, the kind, or the site of a modifying group on the amino acids constituting the histone molecule easily, it would lead to a detection of these diseases in the early stage and facilitate understanding of the diseases.

So far, use of a tandem mass spectrometry (MS/MS) was studied as a method of determining the modification state of amino acid residues (Patent Document 1).

[Patent Document 1] Japanese Laid-open patent publication NO. 2002-100318

DISCLOSURE OF THE INVENTION

However, use of MS/MS demands a mass spectrometer allowing MS/MS measurement. The MS/MS measurement also demanded a longer period than MS measurement. Thus, MS/MS has room for some improvement from the viewpoint of simplicity of the measurement. The fragment ions obtained by MS/MS depend largely on the properties of the peptide used, and it was occasionally difficult to determine the amino acid sequence and the modification state thereof accurately, depending on proteins. For example, it is often difficult to detect phosphorylation sites by methods such as LC/MS/MS. In particular, it was especially difficult to detect phosphorylation sites when the

phosphorylation sites are located in a long peptide or at sites significantly distant from the both terminals of the peptide. In addition, the signal of negatively charged phosphorylated serine is practically very weak in many cases, and thus, MS/MS is not used widely as a common method.

An object of the present invention, which was made in view of the circumstances above, is to provide a method of easily obtaining information about the presence or absence of modification on a particular amino acid residue in a protein and the modifying group together with information about the site of the particular amino acid residue.

According to the present invention, there is provided a method of analyzing a protein, including binding a marker to a particular kind of amino acid residue having an unmodified side chain among the particular kinds of amino acid residues constituting the protein, cleaving the protein at a predetermined site and obtaining peptide fragments, measuring the molecular weights of the peptide fragments, and determining the presence or absence of the marker bound to the particular kind of amino acid residue in the peptide fragments containing the particular amino acid residue and determining the presence or absence of the side chain modification or the kind of the modifying group, by comparing the molecular weight of each peptide fragment calculated from the molecular weights of the amino acid residues constituting the peptide fragment with the molecular weight of each peptide fragment measured in the

measuring the molecular weights.

The term "modification" in the present specification means natural modification of a protein, that is, that an amino acid residue of the protein has a naturally introduced side chain. The term "marker" in the present specification means a substance artificially bound to the side chain of an amino acid residue of a protein.

The term unmodified side chain in the present specification means the side chain of an amino acid residue in a protein in the state that it is not completely modified, that is, that a marker may be bound additionally to the side chain. Even when the side chain of an amino acid residue is bound to a predetermined modifying group, if a marker can be bound to it additionally, the side chain is still unmodified.

The analytical method according to the present invention includes binding a marker to a particular kind of amino acid residue having the unmodified side chain among particular kind of amino acid residues constituting the protein. In this way, it is possible to determine whether a marker is bound to a particular kind of amino acid residue, by comparing the molecular weight of the peptide fragment calculated from the molecular weights of the amino acid residues constituting the peptide fragment with the molecular weight of the peptide fragment measured in the measuring molecular weight. In most cases, the side chain of the particular kind of amino acid residue in the peptide

fragment is apparently unmodified when a marker is bound, while the particular kind of amino acid residue is already modified when the marker is not bound.

By the analytical method according to the present invention, which determines the presence or absence of modification on the side chain of a particular kind of amino acid residue in each peptide fragment or the kind of the modifying group, it is possible to obtain information about the modification site, that is, about which side chain of an amino acid residue contained in which peptide fragment is modified, even when there is a plurality of particular amino acid residues in a protein. Thus by the method according to the present invention, it is possible to perform a protein analysis easily concerning the presence or absence, the number, and the kinds of the modifying groups, as well as the sites of the modifying groups. In the present specification, the state of the side chain of an amino acid residue concerning the presence or absence, the number, and the kinds of the modifying groups, and the sites of the modifying groups will be referred to jointly as its modification state.

Thus, it is possible to obtain information about the modification state of the side chain of an amino acid residue of a protein in an easier manner by the present invention by including binding a marker. It is thus possible to obtain information about posttranslational modification of proteins easily. It is thus possible to obtain information

about various diseases, for example including carcinogenesis, leukemia, aging and the like, by applying the analytical method according to the present invention to the epigenetics field. For example, the method allows, screening of cancers.

The marker is preferably a group that allows identification of the binding of the marker by increase in molecular weight. The molecular weight of the functional group used previously for modifying a particular amino acid residue, such as methyl, acetyl, or ubiquityl group when the particular amino acid residue is a lysine residue, is preferably different from the molecular weight of the marker to be bound. In this manner, by comparing the molecular weight of the peptide fragment calculated from the molecular weights of the amino acid residues constituting the peptide fragment with the molecular weight of the peptide fragment measured in the measuring molecular weight, it is possible to determine easily whether the particular amino acid residue has already a modifying group, whether it is unmodified, that is, whether it still has a marker-binding site.

The analytical method according to the present invention also includes cleaving the protein at a predetermined site and obtaining peptide fragments. The cleavage of the protein at a predetermined site means that the protein is cleaved at a site before or after a certain amino acid residue. Cleavage of the protein at a

predetermined site gives peptide fragments of the protein cleaved according to a certain predictable rule. It is thus possible to calculate the molecular weight of the peptide fragment obtained from the molecular weights of the amino acid residues constituting the peptide fragment. It is also possible to stably obtain information about the site of the modifying group.

The molecular weight calculated from the molecular weights of the amino acid residues constituting the peptide fragment is the molecular weight of the peptide fragment having no modifying group on the side chain. On the other hand, the molecular weight of the peptide fragment measured in the measuring molecular weight is the sum of the calculated molecular weight, the molecular weight of previously modified side chain, and the molecular weight of the marker bound in the binding a marker. In the present invention, it is possible to determine the presence or absence of side chain modification and the kind of the modifying group by comparing these molecular weights.

In the method of analyzing a protein according to the present invention, determining the bound marker, the presence or absence of side chain modification, or the kinds of the modifying groups may include comparing the difference obtained by subtracting the molecular weight of each peptide fragment calculated from the molecular weights of the amino acid residues constituting the peptide fragment from the molecular weights of each peptide fragment measured in the

measuring the molecular weights and with the increase in molecular weight when the marker is bound, and comparing the difference above with the molecular weight of the side chain that can modify the particular kind of amino acid residue.

In the present invention, the comparison with the increase in molecular weight when a marker is bound is to compare with the molecular weight of the marker to be bound to a particular amino acid residue. Alternatively, comparison with the molecular weight of the side chain that can modify the particular kind of amino acid residue is to compare with the molecular weight of the functional group possibly modifying a kind of amino acid, such as methyl, acetyl, or ubiquityl group when the particular amino acid residue is a lysine residue. The presence of these steps makes determination of the presence or absence of side chain modifications or the kind of the modifying group more reliable.

In the method of analyzing a protein according to the present invention, obtaining the peptide fragments may include treating the marker-bound protein with a protease, and the sensitivity of the particular kind of amino acid residue to the protease treatment may be eliminated by the marker.

Presence of the treating protein with a protease ensures cleavage of the protein at a constant site and gives peptide fragments more reliably. Even when the particular

kind of amino acid residue has different sensitivities against a protease depending on the presence or absence of a modifying group, it is possible to prevent fluctuation in sensitivities by eliminating the sensitivity of the bound marker to a protease treatment specific to a particular kind of amino acid residue. It is thus possible to cleave the protein more accurately at a predetermined site. It is thus possible to obtain peptide fragments cleaved at a predetermined site further more reliably.

In the method of analyzing a protein according to the present invention, the obtaining the peptide fragments may contain obtaining at least one peptide fragment containing the particular kind of a single amino acid residue. By obtaining the peptide fragment containing only one particular kind of amino acid residue, it becomes possible to determine the modification state of the particular kind of amino acid residue in the peptide fragment accurately. In the present invention, it is possible to obtain information about the number of residues from N-terminal side of the amino acid residue by including obtaining at least one peptide fragment containing only one particular kind of amino acid residue. It is thus possible to determine the site of the particular amino acid residue and the modification state of the side chain.

In the method of analyzing a protein according to the present invention, the measuring the molecular weights of the peptide fragments may include measuring the mass numbers

of the peptide fragments by mass spectrometry. It is thus possible to measure the molecular weight of the peptide fragment more easily. In the present invention, which demands no MS/MS measurement or the like, it is possible to simplify a series of analytical operations.

In the method of analyzing a protein according to the present invention, the determining the presence or absence of side chain modification or the kinds of the modifying groups may include determining the number of methylation, acetylation, or ubiquitylation of the lysine residues in the peptide fragments or the presence or absence of phosphorylation on the serine or threonine residues in the peptide fragments. It is thus possible to obtain information about the posttranslational modification of the lysine or serine residue in a protein by a simpler method.

In the method of analyzing a protein according to the present invention, the binding a marker may include N-acylating the lysine residues in the protein. It is thus possible to bind a marker molecule to an unmodified lysine residue reliably.

In the method of analyzing a protein according to the present invention, the binding a marker may include succinylating the particular kind of amino acid residue. It is thus possible to bind a marker succinyl group to the unmodified side chain of a particular kind of amino acid residue reliably and ensure favorable water solubility of the marker-bound protein. It is thus possible to advance

the subsequent steps below further more reliably. When the marker binds partially, it is possible to obtain information about partial modification.

In the method of analyzing a protein according to the present invention, the obtaining the peptide fragments may include digesting the marker-bound protein with trypsin. It is thus possible to cleave the protein at the C-terminal side of a basic amino acid residue. It is thus possible to predict the cleavage site from the amino acid sequence of the protein. It is thus possible to determine the presence or absence of side chain modification or the kinds of the modifying groups further more reliably.

In the method of analyzing a protein according to the present invention, the protein may be histone. It is thus possible to analyze the modification state of the histone lysine or serine residue easily. It is thus possible to determine the presence or absence, the kind, the site of the modifying groups on the amino acids constituting the histone molecule and the degree of modification easily. It is thus possible to apply the method favorably to tests of various diseases including cancer screening.

As described above, the present invention provides a method of easily obtaining information about the presence or absence of modification on a particular amino acid residue in a protein and the modifying group together with information about the site of the particular amino acid residue.

BRIEF DESCRIPTION OF THE DRAWINGS

The objects described above, other objects, the characteristics and advantages of the invention will be more apparent with reference to the favorable embodiments described below and the following drawing associated therewith.

Fig. 1 is a flow chart showing the procedure of analyzing a protein in an embodiment.

Fig. 2 is a flow chart showing the procedure of analyzing a protein in another embodiment.

Fig. 3 is a chart showing the analytical results of histone H3 in an Example.

Fig. 4 is a chart showing the analytical results of histone H3 in another Example.

Fig. 5 is a chart showing the analytical results of histone H3 in another Example.

Fig. 6 is a chart showing the analytical results of histone H3 in another Example.

Fig. 7 is a chart showing the analytical results of histone H3 in another Example.

Fig. 8 is a chart showing the sites of modified amino acid residues in histone H3 so far proposed.

BEST MODE FOR CARRYING OUT THE INVENTION

Hereinafter, favorable embodiments of the present invention will be described.

(First embodiment)

The present embodiment relates to a method of determining the modification state of an amino acid-residue side chain in a purified protein. Specifically, the method allows determination of the presence or absence and the number of lysine residues acetylated, methylated, or ubiquitylated, or the presence or absence and the number of arginine residues methylated, in a protein to be measured. It also allows determination of the presence or absence of serine residues phosphorylated.

Fig. 1 is a flow chart showing the procedure for analyzing protein in the present embodiment. First in Fig. 1, a reagent for labeling the side chain of a particular amino acid residue in a protein to be analyzed is allowed to react. In the present specification, the "labeling" of side chain means that a labeling substance is introduced intentionally on the side chain of an amino acid residue of a protein. When the side chain of the particular amino acid residue is in the unmodified state, the side chain is labeled (S101). The protein containing particular amino acid residues having the labeled side chain is then cleaved at a predetermined site specifically (S102). The peptide fragments thus obtained are then analyzed by mass spectrometry (S103). Then, the modification state of the

side chain of the particular amino acid residue in the protein to be analyzed is analyzed (S104), based on the results obtained on the molecular weights of the fragments,

In step 101, the side chain of an amino acid residue of which the modification state is to be analyzed is selectively labeled. The labeling substance for use then is a substance having a certain molecular weight, for example, a predetermined molecule. The reagent is so selected that the molecular weight of the protein-bound labeling molecule differs significantly from the molecular weight of the modifying group to be analyzed. In this way, it is possible to analyze the modification state of the protein reliably in the steps 104 described later, based on the information about the molecular weights of the peptide fragments obtained in step 103.

It is also preferable to select a reagent that causes as small deterioration as possible in the solubility of the protein in water by labeling. It is thus possible to prevent insolubilization and accompanying aggregation of the protein. It is thus possible to make the procedure in the steps 102 and below proceed more reliably.

By labeling the side chain of the particular amino acid residue in step 101, it is possible to reduce the difference in modification sensitivity even when the sensitivity of the particular amino acid residue to the cleavage reagent described below used in the step 102 varies according to the presence or absence of modification on the side chain.

It is thus possible to cleave the protein at a predetermined site reliably, while preventing occurrence of uneven cleavage.

A method of labeling the side chain of a particular amino acid residue is, for example, to N-acylate the side chain of a basic amino acid residue by using an alkanoic anhydride. Examples of the alkanoic acid anhydrides favorably used include symmetric anhydrides of alkanoic acids having carbon atoms of 2 or more and 4 or less. In this way, it is possible to reduce generation of the steric hindrance due to introduction of a bulky labeling group. It is thus possible to perform the labeling in step 101 reliably. It is also possible to advance the processings in the steps 102 and below reliably.

During acylation of the side chain of a basic amino acid residue by using an alkanoic acid anhydride, it is preferable to use a compound that makes the protein favorably water soluble after labeling. In this way, it is possible to advance the processings in the steps 102 described later more reliably than by labeling with a substance that makes the protein water-insoluble after labeling, for example, when acetyl-d3 labeling is performed by using deuterated acetic anhydride. Specifically, it is possible, for example, to succinylate by using succinic anhydride or maleylate by using maleic anhydride.

Among them, for example, succinylation is used favorably. The succinylation results in succinylation of

the side chain of an unmodified amino acid residue, and the succinylated group can be used as the marker of the unmodified group. The succinylation also allows making the protein favorably water soluble after labeling. Because generally the difference in molecular weight between the labeling substance and the side chain-modifying functional group such as acetyl or methyl group is relatively large, it is possible to analyze the modification state more accurately and reliably in the step 104 described below.

The succinylation can be, for example, performed according to the following procedure. First, a protein is dissolved or dispersed in a predetermined solvent. Succinic anhydride is then added thereto to a predetermined concentration, and allowed to react in a basicity condition at predetermined temperature for predetermined period. For prevention of the decrease in the concentration of succinic anhydride associated with the progress of reaction, it is preferable to add succinic anhydride in a great excessive amount with respect to the protein to be reacted. The excess succinic anhydride is then neutralized after reaction.

In step 102, enzyme digestion (enzymatic degradation) by using a protease specific to the cleavage site may be performed during the cleavage of the protein at a specific site. It is also possible to use a method by using a chemical reagent such as CNBr specifically cleaving the C-terminal amide bond of a methionine residue. Hereinafter, the

description will be taking the enzymatic degradation as an example, in the present embodiment.

Enzymatic degradation is performed for cleavage of a protein to be analyzed at a site before or after a particular kind of amino acid residue. The enzymatic degradation gives peptide fragments suitable for mass spectroscopy analysis. It is also possible to predict the cleavage site previously, because the enzyme cleaves a protein having a known amino acid sequence at the site before or after a particular kind of amino acid residue. It is thus possible to predict the molecular weights of the peptide fragments to be obtained previously. It is also possible to reduce the number of particular amino acid residues, for example, the number of lysine residues, contained in a single fragment. It is preferable that a single particular amino acid residue, for example, a single lysine residue is contained in a single fragment. In such a way, it is possible to analyze the modification state of the side chain of the particular amino acid residue reliably.

For example, a serine protease trypsin can be used as the enzyme for use in the enzymatic degradation. When trypsin is used, it can be considered that cleavage is performed at the C-terminal of an arginine or lysine residue because cleavage is generally performed at the C-terminal side of a basic amino acid residue. In the present embodiment, it is possible to stop trypsin digestion with respect to the lysine residue, by previously labeling the

side chain of the unmodified lysine residue in step 101 to protect. It is thus possible to degrade a protein to be measured selectively at the C-terminal side of an arginine residue.

Although arginine residues are often modified, for example, normally by methylation, depending on protein to be analyzed, the modified groups are also degraded by trypsin digestion, and thus, it is possible to obtain peptide fragments cleaved at a predetermined site, independently of the presence or absence of modification on the arginine residues.

Examples of the mass spectrometers used in the mass spectrometry in the step 103 include ion trap mass spectrometer, quadrupole mass spectrometer, magnetic-field mass spectrometer, time-of-flight (TOF) mass spectrometer, Fourier-transform mass spectrometer, and the like. Examples of the ionization methods include electrospray ionization (ESI) method, matrix-assisted laser desorption/ionization (MALDI) method, fast atom collision ionization (FAB) method, and the like.

Among them, MALDI-TOF-MS is used favorably. Use of the MALDI-TOF-MS method prevents missing of the atomic groups in part of the amino acid residues constituting the protein in the ionization process. It is also possible to analyze peptide fragments having relatively higher molecular weight favorably. In addition, even when a protein to be measured in a sample is separated by gel

electrophoresis, treated as above in the gel, then recovered therefrom, and analyzed, it is possible to measure both corresponding anions and cations. Thus, use of the MALDI-TOF-MS method permits more reliable analysis.

The length of the peptide fragments analyzed by mass spectrometry in the step 103 is preferably 20 to 30 amino acid residues or less. In this manner, it is possible to ionize the peptide fragments more reliably during mass spectrometry.

In the step 104, the molecular weight of each of the peptide fragments measured by mass spectrometry is compared with the molecular weight calculated from the primary structure information. If the primary structure of a protein is known, it is possible to predict the site cleaved in the step 102 and thus to calculate the molecular weight of each of the peptide fragments when the side chain has no modifying group.

When the side chain of an amino acid residue to be analyzed is unmodified, there is an increase by the molecular weight obtained by subtracting the molecular weight of a hydrogen atom from the molecular weight of labeling with respect to the molecular weight calculated from primary structure information. Alternatively when the side chain is previously subjected to a predetermined modification, there is an increase by the molecular weight obtained by subtracting the molecular weight of a hydrogen atom from the molecular weight of the modifying group multiplied by

the number of modifying groups. Then, the difference between the molecular weight of each of the peptide fragments measured by mass spectrometry and the molecular weight calculated from primary structure information is determined, and thus, it is possible to determine whether there is modification on the side chain of the amino acid residue to be analyzed present in each peptide fragment and the number of modifying groups by using the difference.

For example, when the side chain of a lysine residue is unmodified and a succinyl group is used as a labeling agent, there is an increase by the molecular weight of the succinyl group 100.07.

It is observed that molecular weight equivalent to the difference in molecular weight between -H and $-CH_3$, that is, CH_2 of 14.03, is added to the molecular weight of the peptide fragment calculated from the amino acid sequences when the side chain of an amino acid residue to be analyzed is methylated because -H is substituted with $-CH_3$; an increase equivalent to the difference between $-COCH_3$ and -H of 42.04, when acetylated; and an increase equivalent to the difference between $-PO_3$ and -H of 79.01, when phosphorylated.

It is thus possible to determine the modification state of the side chains of the amino acid residues in the peptide fragment easily, by comparing the difference between the molecular weight of each of the peptide fragments measured by mass spectrometry and the molecular weight calculated

from primary structure information with the increase in molecular weight when a marker is bound and the increase in molecular weight when a possibly modifying functional group is bound. It is thus possible to analyze the presence or absence of a modifying group on the side chain of an amino acid residue or the kind of the modifying group, based on the mass spectrum obtained in the step 103.

In the present embodiment, it is also possible to determine the modification state of an amino acid residue in each peptide fragment of the protein to be analyzed. Even when there is a plurality of amino acid residues to be analyzed in a protein to be analyzed, it is possible to determine which modifying group is bound to the amino acid residue in which peptide fragment or the amino acid residue in which peptide fragment is labeled. It is thus possible to obtain information about the modification site easily and rapidly. For example, when a single amino acid residue to be analyzed is present in a peptide fragment, it is possible to determine the number of residues from the N-terminal to the amino acid residue. Even when there is a plurality of amino acid residues to be analyzed in a peptide fragment, it is also possible to determine which amino acid residue in the peptide fragment is modified, by reasoning by analogy from other information.

It is possible to analyze the modification state of the side chain of a particular amino acid residue constituting a protein easily by the procedure above. As

a result, it is possible, for example, to obtain information about the presence or absence of posttranslational modification of protein or the kinds of the modifying groups easily, together with information about the modification sites. The protein for use in the analytical method in the present embodiment preferably contains an amino acid residue susceptible to enzymatic degradation. It also preferably contains an amino acid residue to be previously modified such as lysine, serine, threonine, or arginine residue. An example of the protein for use is a basic protein. A typical example of the basic protein is histone.

In the present embodiment, the protein to be analyzed may be previously substituted with another labeling agent for detection for determination of the presence or absence of phosphorylation on a serine or threonine residue. For example, it is possible to obtain an aminoethylcysteine-modified protein, by dephophorylating the protein with $Ba(OH)_2$ and substituting with aminoethylcysteine (NH_2 C_2H_4-SH). When the protein is succinylated by the above-mentioned method, if the protein is phosphorylated, it is observed that molecular weight of 115.8 corresponding to succinylaminoethylcysteine modification is added to the molecular weight of the peptide fragment calculated from the amino acid sequence. At this time, trypsin digestion is not performed. It is possible to intensify the signal of the phosphorylated serine in mass spectrometry by substitution of a modifying group. Thus,

it is possible to detect the presence or absence of phosphorylation at further higher accuracy.

Specifically, 0.125 M $Ba(OH)_2$ is added to a protein or peptide containing a phosphorylated serine or threonine, and the mixture is allowed to react at 40°C for 1 hour, to give a dehydroamino acid residue. For example, phosphorylated serine gives dehydroalanine. The mixture is cooled, and then, small pieces of dry ice are added thereto at low temperature, for white turbidity $Ba(OH)_2$ as barium carbonate and removal thereof. Then, 5-(2-aminomethyl) cysteine is added in an amount at a molar ratio of 500 times as large as that of the protein or peptide. 1 w/v% NH_4HCO_3 is then added thereto, and the mixture is allowed to react at 20°C for 1 hour, to give an aminoethylcysteine derivative. The derivative, which contains an amino group like lysine, may be then processed sequentially in the steps followed step S101 by the above-mentioned method. For example, after succinylation at 40°C for 1 hour, the derivative may be subjected to trypsin degradation at 40°C for 1 hour before analysis by mass spectrometry.

In the present embodiment, the protein to be analyzed may be subjected to a pretreatment of separation and purification before the step 101 shown in Fig. 1. In this way, it is possible to separate a particular protein from a desirable sample and analyze the modification state of the side chain. For example, for analysis of histone H3, a favorable protein sample can be extracted from bovine liver

or the like.

In the present embodiment, if the protein to be measured contains a reduced cysteine residue, the side chain -SH group is preferably protected previously, for example, by carboxymethylation or pyridylethylation. Alternatively, if the protein to be measured contains a cysteine residue having an intramolecular or intermolecular disulfide bond, it is preferable to previously convert the cysteine residue into the reduced form, by a common reduction method. In addition, for prevention of reoxidation of the reduced cysteine residue, it is preferable to keep the reaction system in the step 101 in a condition free from oxygen and moisture. For example, it is possible to conduct the treatment in the step 101 in an inert gas atmosphere such as of nitrogen.

(Second embodiment)

The analytical method described in the first embodiment may include additionally purifying a protein to be analyzed from a sample containing multiple components before the side chain-labeling in the step 101.

Fig. 2 is a flow chart showing the procedure of analyzing a protein in the present embodiment. The procedure shown in Fig. 2 includes a protein-purifying step (S106) before the step 101 in the procedure shown in Fig. 1. The protein-purifying step is, for example, gel electrophoresis. Use of gel electrophoresis allows

reliable fractionation of the proteins in the sample according to their respective molecular weights. The protein to be analyzed fractionated by electrophoresis may be processed in the step 101 while it is contained in the gel. In this way, it is possible to perform a series of processings of separating a protein to be analyzed from a sample and analyzing the modification state of the side chain more easily.

Hereinafter, a method of determining the modifying group will be described, taking a case where the protein to be analyzed is histone as an example. Histones are grouped into types H1 (molecular weight: 21,000), H2A (molecular weight: 14,500), H2B (molecular weight: 13,700), H3 (molecular weight: 15,300), and H4 (molecular weight: 11,300). A nucleosome constituting chromatin contains a histone core consisting of an octamer of 2×H2A, 2×H2B, 2×H3, and 2×H4, and the histones H3 and H4 therein bind to the DNA. In the present description, histones H1, H2A, H2B, H3, and H4 are referred to jointly as "histones".

Posttranslational modification of the amino acid residues of the histone, such as phosphorylation, acetylation, or methylation, is believed to be responsible for the change in chromatin structure and regulation of activation and inactivation of transcription. Accordingly, if the modification state of the lysine, arginine, or serine residue in histone can be determined by a simpler method, it becomes possible to apply the method, for example, to

detection of diseases associated with transcription abnormality and others.

In the present embodiment, a method of obtaining mass spectra of the peptides obtained by previous succinylation and trypsin degradation of bovine histone H3 protein will be described as an example. Histone molecules are first separated from the tissue as follows:

Separation and purification of histones from the tissue can be performed, for example, by the following procedure. First, basic proteins are extracted from the cells or a slice of the tissue. In a case where extraction is performed from whole cell or tissue section, they are usually ground into fine fragments or sliced into a thin sample before extraction.

Alternatively when a sample is prepared from cell nuclei, a cell suspension is placed in a homogenizer, and separation of the nuclei is judged by observation under a microscope. The homogenizer for use can be, for example, Potter glass homogenizer B. It is preferable to use a homogenizer in which the pestle and the glass wall are in loose contact with each other. The suspension for use is, for example, a Tris buffer added by NP-40 (Igepal CA630) and deoxycholic acid. The nuclei may be collected by centrifugation and washed with the above-mentioned buffer. The centrifugation condition may be, for example, 3000 × g, 3 minutes, and 25°C.

6 M urea and 0.3 M HCl are added to the pellet obtained

by centrifugation, and the mixture is blended, for example, in a vortex mixer. The suspension is centrifuged, to give a basic protein extract as the supernatant. The centrifugation condition is, for example, 18,000 rpm, 15 min, and 4°C.

The extract is fractionated by urea aluminum lactate starch gel electrophoresis. For example, a horizontal electrophoresis apparatus is used for electrophoresis. Hydrolyzed starch for electrophoresis is suspended in 0.05 M aluminum lactate buffer solution containing 6 M urea at pH 3.5 so as to be a concentration of 10 w/w %, and the mixture is treated at 70°C for 30 minutes while stirred, to give a semitransparent viscous starch gel. The gel obtained is poured into a tray for electrophoresis apparatus and allowed to harden at room temperature. The size of the tray is, for example, 8 cm × 15 cm × 0.5 cm.

The sample is placed in a 80 μL sample slot on the gel while it is acidic, and subjected to electrophoresis. Filter papers connected to both ends of the gel is immersed into reservoirs filled with 6 M urea aluminum lactate buffer solution. Electrophoresis is performed at a constant voltage of 100 V at 4°C for 12 hours. It is terminated when methyl green used as the indicator reaches close to the end of the cathode of the gel.

Staining of the gel is performed then as follows: The gel is first stained with an acetic acid methanol solution of Amido Black, and the stained gel is washed with the same

solution containing no dye several times. Fractionation with 0.5 M sulfuric acid after the Amido Black staining by the method above gives arginine-containing proteins stained black and proteins higher in the lysine and histidine contents that are stained blue. In addition, proteins containing a smaller amount of basic amino acids are stained white and transparent by negative staining. Destaining with an acetic acid methanol solution after the fractional staining visualizes all proteins as bluish bands. Histones are located close to the intermediate point between the methyl green indicator and the electrophoresis starting point, and a fraction containing histones H2A and H2B, a fraction containing H3 and a fraction containing H4 are obtained as separated from each other.

The histones fractionated in the gel can be purified according to the following method. First, each band of histone is cut, homogenized, and separated by gel filtration in an acrylamide gel column by using 0.1 M HCl. It is also purified in a cellulose phosphate column. It may be purified by elution with 6 M urea with sodium chloride concentration gradient under the presence of 6 M urea. Alternatively, the histones may be separated by subjecting the starch gel after electrophoresis additionally to another electrophoresis, eluting the histones therefrom, and collecting them by using a molecular sieve membrane.

Each of the histone fragments thus obtained is then processed in the steps 101 to 104 by using the method

described in the first embodiment. The fractions to be analyzed are processed further in the succinylation reaction, to make the unmodified lysine in the histone molecule succinylated.

The amino acid sequence of calf histone H3 is shown as the sequence No. 1 in sequence table. Table 1 also shows peptide fragments obtained by trypsin digestion of calf histone H3 together with the number of amino acid residues from N terminal. The sequence number in the sequence table for each of the fractions containing four or more amino acid residues among fractions 1 to 16 is also shown in Table 1. As apparent from Table 1, trypsin digestion generates peptide fragments 1 to 16. However, because there are two successive arginine residues as the 52nd and 53rd amino acids, there is a possibility of generating a fragment 8' partially as well as fragment 8.

As apparent from Table 1, it is possible to decompose a protein having many lysine residues at the N-terminal chain such as histone H3 into peptide fragments containing one to three lysine residues, by trypsin digestion after succinylation. It is thus possible to analyze acetylation, methylation, or ubiquitylation of the lysine or arginine residue or phosphorylation of the serine or threonine residue in each fragment easily by the method described in the first embodiment.

TABLE. 1

SEQUENCE NUMBER	FRAGMENT NUMBER	AMINO ACID RESIDUE NUMBER	AMINO ACID SEQUENCE	MOLECULAR WEIGHT
–	1	1–2	AlaArg	245.159
2	2	3–8	ThrLysGlnThrAlaArg	703.398
3	3	9–17	LysSerThrGlyGlyLysAlaProArg	900.514
4	4	18–26	LysGlnLeuAlaThrLysAlaAlaArg	985.603
5	5	27–40	LysSerAlaProAlaThrGlyGlyValLysLysProHisArg	1432.826
6	6	41–49	TyrArgProGlyThrValAlaLeuArg	1031.588
–	7	50–52	GluIleArg	416.238
7	8	53–63	ArgTyrAlaLysSerThrGluLeuLeuIleArg	1348.793
8	8'	54–63	TyrAlaLysSerThrGluLeuLeuIleArg	1192.681
9	9	64–69	LysLeuProPheGlnArg	787.470
–	10	70–72	LeuValArg	386.264
10	11	73–83	GluIleAlaGlnAspPheLysThrAspLeuArg	1334.683
11	12	84–116	PheGlnSerSerAlaValMetAlaLeuGlnGluAlaSerGluAlaTyrLeu ValGlyLeuPheGluAspThrAsnLeuCysAlaIleHisAlaLysArg	3611.775
12	13	117–128	ValThrIleMetProLysAspIleGlnLeuAlaArg	1383.791
–	14	129–131	ArgIleArg	447.572
–	15	132–134	GlyGluArg	361.412
–	16	135	Ala	89.091

In the present embodiment, extraction and purification of histones may be performed as follows: HCl and LiCl whose final concentrations are 0.165 M respectively are added when the histone-containing tissue is ground, and the supernatant is developed by one-dimensional polyacrylamide electrophoresis (1DE). Each histone band obtained by electrophoresis is separated, and dehydrated repeatedly with acetonitrile (MeCN). Histone extraction from tissues has so far been, for example, performed with HCl at a final concentration of 0.33 M, and thus, the protein is treated with a relatively strong acid. In contrast, in the method of the present embodiment, which allows extraction in an acid containing a sufficient amount of chloride ions and additionally at a low concentration, it is possible to extract histones reliably under a mild

condition.

Then, a formamide (FA) solution of 500 folds succinic anhydride (SucAnh) by mole is added to the dehydrated gel sample, and the mixture is subjected to highly water-soluble succinylation (Suc) at 40°C for 1 hour. This gives rise to an increase in molecular weight of 100.1 per succinyl group. The succinylation is preferably performed as rapidly as possible, for prevention of adverse affect due to the deacetylation in the presence of acetyl groups by acetyl lyases.

Histone, a protein extremely rich in lysine and arginine residues, gives its lysine residues not in the unmodified state, because all of the side chains of lysine residues are modified for example by acetylation or by labeling with an artificial introduction of a marker, for example, by succinylation. Accordingly, use of a protease such as trypsin then does not result in cleavage at the C-terminal side of lysine residue but cleavage only at the C-terminal side of arginine residue. As described above, the same is true when the arginine residue is converted to methyl arginine by modification with methyl group, and thus, the arginine residue is cleaved in such a case.

The histone after succinylation is then treated with trypsin. The treatment is performed, for example, at room temperature in 50 mM NH_4HCO_3 for 20 minutes. The trypsin treatment gives rise to peptide fragments having arginine at the C terminal except the C-terminal peptide. The

degradation product is then analyzed by mass spectrometry using the method described in the first embodiment, and the molecular weights of the fragment peptides and modifying groups can be detected.

Acetyl groups in histone are unstable and easily deacetylated, for example, by histone acetyltransferase (HAT) contained in the living body. It is possible to suppress deacetylation, for example, by extracting histone at a temperature of approximately 5 to 20°C in the presence of an inhibitor to deacetylation enzymes such as a sodium butyrate solution at a concentration of approximately 10 mg/ml.

The present invention is so far described with reference to the embodiments above. These embodiments are only examples of the present invention, and it should be understood for those skilled in the art that various modifications of the present invention are possible and these modifications are also included in the scope of the present invention.

For example, the method described above is not limited to the cases where the substance to be analyzed is a protein, and may be used widely as an easier method of determining the presence or absence of the modification of a particular amino acid residue in a peptide or the kinds of the modifying groups, together with the information about the sites of the amino acid residue.

Hereinafter, the present invention will be described

in more detail with reference to Examples, but it should be understood that the present invention is not restricted thereby.

(Example 1)

In the present Example, the modifying groups on a commercial available histone H3 were analyzed. The histone H3 used was calf thymus histone H3 (purified preparation) manufactured by F. Hoffmann - La Roche, ltd.

100 μL of an ethanol solution containing 2 mg of succinic anhydride and 2 μL of 1 M NH_4HCO_3 were added to 1 μL of an aqueous 1 v/v % formic acid solution containing 0.5 nmol histone H3, and the mixture was allowed to react under an alkaline condition at 40°C, for 2 hours. The reaction product was dried under vacuum; 1 mL of acetonitrile was added thereto; and excess succinic anhydride was removed three times.

10 μL of 0.5 M NH_4HCO_3 was then added, and the mixture was allowed to react at 40°C for 1 hour. Then, 2 μL of an aqueous trypsin solution was added and allowed to react at 40°C for 1 hour. The histone H3 degradation peptide was treated with Zip-tip. The samples 1 and 2 thus obtained were analyzed by mass spectrometry. The samples 1 and 2 were obtained separately by the method above. A mass spectrometer MALDI TOF MAS Voyager DE RP manufactured by Applied Biosystems was used for mass spectrometry. Figs. 3 and 4 show mass spectra obtained respectively from the

samples 1 and 2.

Additionally, calf thymus histone H3 manufactured by F. Hoffmann - La Roche, ltd. (purified preparation) was subjected to SDS-PAGE, and treated in the polyacrylamide gel as follows, to give a sample.

First, a gel band containing 0.5 nmol of histone H3 was separated, cut to a slice of 1 × 1 mm square, and placed in an Eppendorf tube. 1 mL of water was added thereto; the mixture was stirred for 10 minutes; after removal of water, 1 mL of acetonitrile was added thereto; and the mixture was left at rest for 10 minutes; and acetonitrile was removed. The operation was repeated three times for dehydrating the gel. 10 μL of a formamide solution containing 10 mg/mL succinic anhydride was added to the dehydrated gel slice. Then, 100 μL of formamide and 1 μL of pyridine were added thereto, and the mixture was allowed to react at 40°C for 1.5 hours. After reaction, reaction solution was removed; 1 mL of 50 mM of NH_4HCO_3 was added; the mixture was allowed to react at 40°C for 10 minutes; and the operation above of removing excess succinic anhydride was repeated twice.

The reaction product is then subjected to trypsin treatment according to the above-mentioned method and treated with Zip-tip C_{18}, and the samples 3 and 4 thus obtained were analyzed by mass spectrometry. The samples 3 and 4 were obtained separately by the method above. Figs. 5 and 6 show mass spectra obtained respectively from the samples 3 and 4.

The modification state of the lysine and serine residues was analyzed by comparing the mass spectra shown in Figs. 3 to 6 with the molecular weight of the peptide fragments of histone H3 respectively shown in Table 1. Results are summarized in Table 2.

Each "fragment number" in Table 2 corresponds to each "fragment number" in Table 1. The "molecular weight (unmodified, calculated value)" is a calculated value of the molecular weight of the peptide fragment calculated from the molecular weights of the amino acid residues in each fragment. The calculated value is calculated, assuming that no lysine or serine residue is modified at all. The "molecular weight (measured value)" represents a peak mass (m/z) in the mass spectrum of each sample.

In Table 2, the difference between the "molecular weight (measured value)" and the "molecular weight (without modifying groups and labeling, calculated value)" represents a measured result of the increase in molecular weight due to presence of modifying group or succinylation. The increase is compared with the increase in molecular weight of 100.07 by succinylation, the increase in molecular weight of 14.03 by methylation, the increase in molecular weight of 42.04 by acetylation, and that of 79.01 by phosphorylation. For example, by assuming that: [molecular weight (measured value))-(molecular weight (without modifying groups and labeling, calculated value)] = $100.07x + 14.03y + 42.04z + 79.01w$, most probable values

of x, y, z, and w are calculated. x, y, z, and w each represent an integer of 0 or more.

By using the difference between the "molecular weight (measured value)" and the "molecular weight (without modifying groups and labeling, calculated value)", the kinds and the number of the modifying groups on the lysine, arginine, or serine residue in each peptide fragment are determined as the "modification or label state" in Table 2. The "molecular weight (with modifying groups and labeling, calculated value)" is a calculated value of the molecular weight of each of the peptide fragments in the determined modification state. The difference between the "molecular weight (with modifying and labeling groups, calculated value)" and the "molecular weight (without modifying groups and labeling, calculated value)" is a calculated value of the increase in molecular weight due to presence of a modifying group or succinylation.

In the column of the "modification state" in Table 2, "M" indicates methylation of the lysine or arginine residue, and its subscript number represents the number of methyl groups bound to a single lysine residue. Specifically, for example, "S" in fragment 2 indicates succinylation of a lysine residue. Alternatively, "P" represents phosphorylation of a serine residue. The parenthesized number after the alphabet represents the sequence number of the amino acid having a modifying group, that is, the number of amino acids from the N-terminal. The number

corresponds to the amino acid number shown in Table 1. When multiple numbers are divided as in " , " in the parenthesis, these numbers represent the numbers of amino acids having respective modifying groups. Specifically, for example, the "MPMS (27, 28, 36, 37)" in fragment 5 indicates that the 27th lysine has a single methyl group, the 28th serine has a single phosphate group, the 36th lysine has a single methyl group, and additionally the 37the lysine has a single succinyl group bound. Alternatively, when the multiple numbers are divided by "/" in the parenthesis, the numbers mean that one of the amino acid residues at the amino acid numbers before and after "/" has the modifying group. Specifically, for example, "S(9/14)" indicates that the 9th lysine or the 14th lysine has a single succinyl group bound .

As shown in Table 2, in samples 1 to 4, the kinds and the number of the modifying groups on the peptide fragments corresponding to the fragments 2, 3, 4, 5, 6, 8, 9, 11, and 13 were identified. Trypsin-derived peptide fragments were also detected. When multiple modification states are detected in one fragment in Table 2, it means that the modification states of all histone H3 molecules in the sample are not the same.

When there are multiple lysine residues as in fragment 5, the modification state is shown with reference to Fig. 8. Fig. 8 shows the sites of amino acid residues in histone H3 conventionally proposed as modified. Fig. 8 summarizes the information described in such as Thomas Jenuwein and

C. David Allis, "Translating the Histone Code", Science, 2001, Vol. 293 (Figs. 2 and 3), Eric J. Richards and Sarah C. R. Elgin, "Epigenetic Codes for Heterochromatin Formation and Silencing: Rounding up the Usual Suspects", Cell, 2002, Vol. 108, p. 491 (Fig. 2).

Each amino acid residue is expressed by one character in Fig. 8. "M" above each residue indicates that the site is proposed to be methylated. Similarly, the character "A" above indicates that the side chain is proposed to be acetylated, and the character "P" indicates that the side chain is proposed to be phosphorylated.

As apparent by referring to Fig. 8, in the line of the fragment 5 having a value of "molecular weight (with modifying groups and labeling, calculated value)" of 1,640.0 in Table 2, the modification state is mainly "MPMS (27, 28, 36, 37)". Lysine is present at the 14th site from the N terminal and arginine is present at the 17th site in the fragment 3; but, as apparent from Fig. 8, in the line of a value of "molecular weight (with modifying groups and labeling, calculated value)" at 1114.7, the modification state is mainly "MSM (9, 14, 17)". Alternatively, in the line of a value of "molecular weight (with modifying groups and labeling, calculated value)" at 1,014.6, the modification state may be regarded as "MS (9, 14)".

As described above, in the present Example, with all samples succinylated in aqueous solution or gel, it was possible to obtain mass spectra reflecting the modification

states of the lysine and serine residues, as in Figs. 3 to 6. In addition, as shown in Table 2, it was also possible to analyze the presence or absence and the kinds of the modifying groups on part of lysine and serine residues reproducibly. Thus in the Example, it was possible to determine the modifying groups on the side chain of histone H3 by an easier method.

(Example 2)

TABLE. 2

FRAGMENT NUMBER *1)	MOLECULAR WEIGHT (WITHOUT MODIFYING GROUP AND LABELING, CALCULATED VALUE)	MOLECULAR WEIGHT (MEASURED VALUE)					MODIFICATION OR LABELING STATE *2)	MOLECULAR WEIGHT (WITH MODIFYING GROUP AND LABELING, CALCULATED VALUE)
		SAMPLE 1	SAMPLE 2	SAMPLE 3	SAMPLE 4	SAMPLE 5		
2	703.4	744.1	-	-	745.3	-	M_3 (4)	745.4
2	703.4	804.5	-	-	806.0	804.3	S (4)	803.5
3	900.5	1001.5	1002.2	1001.5	1002.8	999.3	S (9/14)	1000.6
3	900.5	-	-	1101.6	-	-	SS (9,14)	1100.7
3	900.6	1115.5	1116.5	1115.6	1116.7	-	MSM (9,14,17)	1114.7
4	985.6	1187.0	1187.6	1186.7	1188.2	1183.7	SS (18,23)	1185.7
5	1432.8	1642.7	1642.6	-	1664.4	1641.1	MPMS (27,28,36,37)	1640.0
5	1432.8	1735.5	1735.6	1733.9	1736.6	-	SSS (27,36,37)	1733.0
6	1031.6	1032.8	1033.4	1032.6	1033.8	1030.4	NONE	1031.6
8	1348.8	-	1448.9	-	1452.3	1449.3	S (56)	1448.9
9	787.5	888.7	889.2	888.5	889.5	886.5	S (64)	887.5
11	1334.7	1435.6	1436.8	1435.7	1432.9	1432.9	S (79)	1434.8
13	1383.8	1485.6	1486.2	1484.8	1487.1	1483.9	S (122)	1483.9
Try	-	841.7	843.2	842.5	842.4	-	-	843.0
Try	-	910.7	911.3	-	-	-	-	908.5

*1) Try : PEPTIDE FRAGMENT DERIVED FROM TRYPSIN

*2) S : SUCCINYLATION OF LYSINE RESIDUE

(N): AMINO ACID NUMBER FROM N TERMINAL

(9/14): 9 OR 14

(9,14): 9 AND 14

M_n : METHYLATION OF LYSINE OR ARGININE RESIDUE
(n : NUMBER OF METHYL GROUPS)

A: ACETYLATION OF LYSINE RESIDUE

P : PHOSPHORYLATION OF SERINE RESIDUE

In the present Example, histone H3 was extracted from bovine liver, and the modifying groups thereon were analyzed. 10 to 50 mg of bovine liver stored frozen at -85°C was placed in a micro Potter glass homogenizer manufactured by Wheaton; PMSF (phenylmethanesulfonyl fluoride) at a final concentration of 0.3 mM, HCl at a final concentration of 0.165 M and LiCl at a final concentration 0.165 M were added; and the mixture was homogenized. The mixture was then centrifuged at 13,500 rpm for 30 minutes, and, after removal of the precipitate, the supernatant was dried under vacuum. The dried product thus obtained was fractionated by SDS-PAGE.

A gel band containing 0.5 nmol of histone H3 was separated, cut to a slice of 1 × 1 mm square, and placed in an Eppendorf tube. 1 mL of water was added thereto; the mixture was stirred for 10 minutes; after removal of water, 1 mL of acetonitrile was added thereto; and the mixture was left at rest for 10 minutes; and acetonitrile was removed. The operation was repeated three times for dehydrating the gel. 10 μL of a formamide solution containing 10 mg/mL succinic anhydride was added to the dehydrated gel slice. Then, 100 μL of formamide and 1 μL of pyridine were added thereto, and the mixture was allowed to react at 40°C for 1.5 hours. After reaction, reaction solution was removed; 1 mL of 50 mM of NH_4 HCO_3 was added; the mixture was allowed to react at 40°C for 10 minutes; and the operation above of removing excess succinic anhydride was repeated twice.

The reaction product is then subjected to trypsin treatment and treated with Zip-tip C_{18}, and the sample 5 obtained was analyzed by mass spectrometry by the method described in Example 1. Fig. 7 is a chart showing the mass spectrum obtained.

By comparing the mass spectrum shown in Fig. 7 with the molecular weights of the peptide fragments derived from histone H3 shown in Table 1, the modification state of the lysine and serine residues was analyzed in a similar manner to Example 1. Results are summarized in Table 2.

In the present Example, it was possible to extract histone H3 under a mild condition reliably, by using HCl at a final concentration of 0.165 M and LiCl at a final concentration of 0.165 M for homogenization during extraction of histone H3. As shown in Fig. 7, with the histone H3 extracted from bovine liver, it was possible to obtain mass spectra of the peptide fragments reflecting the modification states of the lysine and serine residues. As shown in Table 2, it was also possible to determine the modification states of part of the lysine and serine residues by a simpler method.